Maastricht University

Learning from big imaging data to predict radiotherapy treatment outcomes and side-effects

Citation for published version (APA):

Shi, Z. (2020). *Learning from big imaging data to predict radiotherapy treatment outcomes and side-effects*. [Doctoral Thesis, Maastricht University]. ProefschriftMaken. https://doi.org/10.26481/dis.20200908zs

Document status and date:
Published: 01/01/2020

DOI:
10.26481/dis.20200908zs

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 05 May. 2023

Appendix II

Valorization addendum

The prevalence of cancer is an increasing healthcare issue as it is the predominant cause of death worldwide. The growing cancer burden is caused by several factors including population growth, aging, and the changing prevalence of certain causes of cancer during social and economic development [1]. To address the global cancer burden, new technologies, for instance Artificial Intelligence (AI), have been applied in the workflow of cancer care from diagnosis to treatment. For cancer treatment, especially radiotherapy, new innovations are not only useful to provide comprehensive treatment plans, but also able to reduce radiotherapy-induced side-effects which may exist in patients during and (long) after treatment.

The clinical data science (CDS) research group of Maastricht University performs data science with the aim to provide clinical decision aid systems for individualized radiotherapy by the following approaches:

1. Developing global FAIR [2] data sharing infrastructures.
2. Learning personalized prediction models from FAIR data.
3. Applying clinical decision aid systems to improve cancer care.

As a CDS researcher, the work of the thesis has contributed to the investigation of big imaging data and new AI technologies for the first two research aims of CDS.

Knowledge dissemination

Medical imaging represents one of the largest segments of big cancer data and remains hitherto an under-utilized data resource. As data is the basis of machine learning algorithms, it is necessary to figure out the way patient data is collected and stored in hospitals. **Chapter 2** introduced the big radiation oncology data from different angles including data collection from different departments, storage in different and formats and systems, as well as the challenges and opportunities for data exchange and machine-assisted learning.

The success of modern oncology implies that some patients live long with the adverse outcomes of their treatment. The knowledge described in **Chapter 3** might be used to guide cancer treatment, so that the quality of patients' lives can be improved after treatment.

Radiomics on multiple modalities is still under study. The previous findings are that the models developed using CT or PET radiomics indeed had prognostic ability. However, as shown in **Chapter 5 and 6**, they may fail in independently external validation, because of different population between training and validation cohorts, inappropriate feature selection in the phase of model training, and so on. Therefore, appropriate validation is a necessary step to assess the generalization of prediction models based on radiomics.

In order to use big imaging cancer data easily and efficiently, the application of FAIR imaging data was described in this thesis. Furthermore, by integrating distributed learning and FAIR data, **Chapter 7 and 8** provided a picture about how machine learning can be implemented in multiple centers without sharing data for the aim of data privacy preserving. The introduced studies can be seen as templates for future distributed radiomics studies in the domain.

Societal or commercial relevance

FAIR data is a novel solution to allow machine algorithms can understand the data well enough to process it automatically with limited human intervention. **Chapter 4** introduced an ontology-guided radiomics analysis workflow (O-RAW) [3] that is able to generate FAIR radiomics data, so that multi-center radiomics research is easier. In addition to knowledge sharing, the O-RAW package has been widely used in the CDS group as it provides a pipeline to use DICOM images as the input and produce RDF FAIR data as the output. Furthermore, the cost of data management, retrieval, and interpretation can be reduced by making cancer data FAIR. It is a potential solution to handle the growing big data in the cancer domain.

Distributed learning has shown the feasibility and importance to allow machine learning algorithms on physically federated data sources. Regarding to commercial applications of distributed learning, the reality is that medical companies are often not allowed to access directly the patient data stored in hospitals, because of data privacy and security regulations such as GDPR [4] and HAAP [5]. The work in **Chapter 7 and 8** provided two distributed learning applications, which have shown a potential solution to medical companies by using the concept of the personal health train infrastructure [6] to implement distributed machine learning.

Although the advance of AI-based medical applications is a long and pricy process, the global market is getting larger and larger. This is reflected by the evidence that more and more companies participated in the industry of AI-based medical imaging in the last 5 years including leading enterprises, such as Google, Microsoft, Facebook, Alibaba, Philips, and so on. Due to the phenomena that AI algorithms and computational power are trending to similarity in the market, the properties of data (e.g., volume and quality) seem to be important factors leading to commercial success. Therefore, the study of big cancer data is not only able to improve work efficiency for hospitals and better cancer for individual patient, but to support the advance of AI-medical applications for medical companies and research institutes.

In the current change of era from internet technology (IT) to data technology (DT), new digital technology is developing dramatically. The new digital technology includes Internet of Things, Cloud Computing, Big Data Technology, AI, and Blockchain. These five blocks are building a framework that might work horizontally as follows: (1) Internet of Things can collect real-time data using multi-sensors and transmit this data to the cloud; (2) The cloud can provide data storage space and computing power; (3) big data technology can manage and curate data; (4) AI can process big data, and extract valuable knowledge and information; and (5) Blockchain can provide security support for the knowledge and information in further use. The framework is expected to work in the healthcare domain as well. Hence, there will be many opportunities for medical enterprises to exploit the healthcare market by using these new technologies.

In conclusion, big imaging data-based prediction models developed by a distributed learning approach have clear valorization potential in cancer care.

References

1. World Health Organization. "*Latest global cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018*." International Agency for Research on Cancer. Geneva: World Health Organization (2018).
2. Wilkinson, Mark D., et al. "*The FAIR Guiding Principles for scientific data management and stewardship*." Scientific data 3.1 (2016): 1-9.
3. O-RAW at: https://gitlab.com/UM-CDS/o-raw
4. Annas, George J. "*HIPAA regulations—a new era of medical-record privacy?."* (2003): 1486-1490.
5. Voigt, Paul, and Axel Von dem Bussche. "*The eu general data protection regulation (gdpr)*." A Practical Guide, 1st Ed., Cham: Springer International Publishing (2017).
6. Personal Health Train at: https://www.health-ri.nl/initiatives/personal-health-train